

Raising the Bar of AI-generated Image Detection with CLIP

Year: 2023 | Citations: 128 | Authors: D. Cozzolino, G. Poggi, Riccardo Corvi

Abstract

The aim of this work is to explore the potential of pre-trained vision-language models (VLMs) for universal detection of AI-generated images. We develop a lightweight detection strategy based on CLIP features and study its performance in a wide variety of challenging scenarios. We find that, contrary to previous beliefs, it is neither necessary nor convenient to use a large domain-specific dataset for training. On the contrary, by using only a handful of example images from a single generative model, a CLIP-based detector exhibits surprising generalization ability and high robustness across different architectures, including recent commercial tools such as Dalle-3, Midjourney v5, and Firefly. We match the state-of-the-art (SoTA) on in-distribution data and significantly improve upon it in terms of generalization to out-of-distribution data (+6% AUC) and robustness to impaired/lauded data (+13%). Our project is available at <https://grip-unina.github.io/ClipBased-SyntheticImageDetection/>